

Article

## Forecasting Regional Sugarcane Yield Based on Time Integral and Spatial Aggregation of MODIS NDVI

Betty Mulianga<sup>1,2,3,\*</sup>, Agnès Bégué<sup>2</sup>, Margareth Simoes<sup>2,4</sup> and Pierre Todoroff<sup>5</sup>

<sup>1</sup> Kenya Sugar Research Foundation (KESREF), Kisumu-Miwani Road, P.O Box 44, Kisumu 40100, Kenya

<sup>2</sup> CIRAD-UMR TETIS, Maison de la Technologie, 500 rue J.-F. Breton, F-34093 Montpellier, France; E-Mail: agnes.begue@cirad.fr

<sup>3</sup> CIRAD-UPR SCA, Av. Agropolis, F-34398 Montpellier Cedex 5, France

<sup>4</sup> EMBRAPA-Programa LabEx Europa and Rio de Janeiro State University PPGMA/DESC/UERJ, Agropolis International, Av. Agropolis, F-34094 Montpellier, France; E-Mail: margareth.simoes@embrapa.br

<sup>5</sup> CIRAD UPR SCA, Station de Ligne-Paradis, 7 chemin de l'IRAT, Saint-Pierre, F-97410 Réunion, France; E-Mail: pierre.todoroff@cirad.fr

\* Author to whom correspondence should be addressed; E-Mail: bmulianga@gmail.com; Tel.: +254-020-204-7307; Fax: +254-020-204-7308.

Received: 15 March 2013; in revised form: 26 April 2013 / Accepted: 26 April 2013 /

Published: 10 May 2013

---

**Abstract:** This study explored the suitability of the Normalized Difference Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectrometer (MODIS) obtained for six sugar management zones, over nine years (2002–2010), to forecast sugarcane yield on an annual and zonal base. To take into account the characteristics of the sugarcane crop management (15-month cycle for a ratoon, accompanied with continuous harvest in Western Kenya), the temporal series of NDVI was normalized through an original weighting method that considered the growth period of the sugarcane crop (wNDVI), and correlated it with historical yield datasets. Results when using wNDVI were consistent with historical yield and significant at P-value = 0.001, while results when using traditional annual NDVI integrated over the calendar year were not significant. This correlation between yield and wNDVI is mainly drawn by the spatial dimension of the data set ( $R^2 = 0.53$ , when all years are aggregated together), rather than by the temporal dimension of the data set ( $R^2 = 0.1$ , when all zones are aggregated). A test on 2012 yield estimation with this model realized a RMSE less than  $5 \text{ t ha}^{-1}$ . Despite progress in the methodology through the weighted

NDVI, and an extensive spatio-temporal analysis, this paper shows the difficulty in forecasting sugarcane yield on an annual base using current satellite low-resolution data. This is particularly true in the context of small scale farmers with fields measuring less than the size of MODIS 250 m pixel, and in the context of a 15-month crop cycle with no seasonal cropping calendar. Future satellite missions should permit monitoring of sugarcane yields using image resolutions that facilitate extraction of crop phenology from a group of individual plots.

**Keywords:** MODIS; NDVI; environment; sugarcane; yield forecasting

---

## 1. Introduction

Sugarcane (*Saccharum* Spp. Hybrids) is a gramineae of the tribe of Andropogonae and Poaceae family. It is defined as a tropical semi perennial crop which is harvested through the manual system in Kenya at variable periods depending on the date of planting, variety, tiny climatic variations along the year and mill preparedness. Consequently, re-growth of sugarcane, known as ratooning, matures at different periods, introducing spatio-temporal variability in the sugarcane landscape. In Kenya where sugarcane is rain fed, this variability is exacerbated with an unspecified cropping calendar and diversification of the cropping system both at spatial (sugar management zone) and temporal (inter-annual) levels, presenting a heterogeneous sugarcane landscape. Sugarcane being the second largest contributor to Kenya's agricultural growth saves the country in excess of USD 229,885,057 annually in foreign exchange, while contributing to poverty reduction and national development [1]. Knowledge of crop productivity is therefore necessary for proper, foresighted and informed planning for competitiveness in the sugar industry [2] and national development.

In Kenya, sugarcane yield is estimated using conventional approaches through biennial field surveys by millers and the Sugar Board, basing their methodology on visual physical assessment (VPA) [3]. In VPA, a stratified random sampling approach is used, considering 15% field coverage in each administrative sector of the zones. A monthly productivity index ranging between 0 and 5 is then applied to sample cane crop from the age of one month, while considering the parameters: (i) crop vigour, (ii) crop colour, (iii) crop density, (iv) weed status, pests and diseases at the time of yield assessment. The average scores are then computed against preset reference yields for each crop cycle with the assumption that the crop has been managed under recommended standard guidelines [1]. The estimated yield is used by the Sugar Industry to project sugarcane production for the current and subsequent year. Although this method has been used since sugarcane was first grown in Kenya, accuracy of manual methods has been proven to introduce gross errors in the results due to variability in time scale and fatigue [4]. This manual method assumes that the crop properties remain constant at the age of yield estimation till crop maturity at 14 to 16 months for ratoons, and 18 to 20 months for plant crop, respectively. Further, it is assumed that 15% of the sample is sufficient to represent crop conditions in the entire mill zone. This could only be true if the crop calendar is defined and not in a spatially heterogeneous landscape such as is the case in Kenya. Similarly, the method assumes that environmental variables such as rainfall distribution and amount will not change in the subsequent

year. More so, the human eye is limited in its ability to discriminate colors of an object quantitatively, compared to multispectral systems [5]. Additionally, physical ground data collection has been proven to be time consuming and unreliable in its temporal scale [6,7]. It is the subjectivity of the current traditional method for monitoring sugarcane production that creates most of the gap for a near real time method that will integrate timely environmental variables in estimating sugarcane yield through a remote sensing approach [4].

Remote sensing is the near real time method. The advantage of remote sensing over ground systems, such as that used by the millers, is that they cover wide areas explicitly, providing timely spatial and temporal data. Such temporal data has been commended for monitoring vegetation development in response to changes in the environment and in response to human management practices [7–10]. These conditions vary over large areas due to diverse topography, soil type, rainfall distribution and management practices, to which sugarcane phenology and productivity is dependent [11]. Most vegetation indices have proven successful in estimating crop yield and biomass [4]. The Normalized Difference Vegetation Index (NDVI) from remote sensing imagery for example, has been expansively used to determine crop phenology, biomass and productivity in spatial distribution [12,13]. The quality of methods developed depends on the scale of study and on the crop management practices, which influence the temporal and spatial resolutions of the relevant data. The cost of satellite imagery, however, is high when fine resolution is required. Crop monitoring studies have therefore resolved this impasse by successfully using free low resolution images from the Moderate Resolution Imaging Spectroradiometer (MODIS), SPOT-VEGETATION, or NOAA-AVHRR sensor data for crop studies [14].

Recent studies have used low resolution imagery to estimate sugarcane yield production in different countries. In Brazil for example [15], 1 km SPOT-VEGETATION data was used, taking advantage of its daily temporal resolution and coupling it with meteorological data to monitor sugarcane development. Cropping seasons were successfully identified using the NDVI data and further facilitated classification of the data for analysis. In the three yield classes assessed ( $24\text{--}73\text{ t ha}^{-1}$ ;  $42\text{--}110\text{ t ha}^{-1}$ , and  $74\text{--}85\text{ t ha}^{-1}$ ), the yield predicted was consistent with the historical yield with accuracies of 8.3%, 66.7% and 86.5%, respectively. The low accuracy of the first class would be attributed to coarseness of the 1 km image that limits discrimination of individual phenology for plots that are smaller than the pixel size, a case similar to the small scale sugarcane farming community of Kenya. Accuracies for the second and third class were in the municipality areas, characterized with large farms such as the nucleus fields of Kenyan sugar mills that are under pure sugarcane stand. A similar study, [11] noted that neither average rainfall nor average MODIS NDVI was related to the average sugarcane yield of the farmers' fields situated within the 5 km radius of the nine weather stations. On a larger scale, MODIS NDVI had a positive correlation ( $R = 0.57$ ) with yield when averaged across all nine management zones, but only for the rainy-season planting. In a different study [16], NOAA-AVHRR data was utilized to develop and validate a model for forecasting crop yield in Pakistan. District data was then used to validate the model, resulting in a root mean square error of  $13.5\text{ t ha}^{-1}$  for sugarcane yield. In their recommendations, actual daily sunshine hours, air temperature, and a crop map were argued to be indispensable for refinement of the model.

A recent study on forecasting sugarcane crop season in Brazil using simple correlations between time series NDVI from AVHRR and an agro-climatic index on sugarcane yield, realized significant

correlations ( $R = 0.69$  to  $0.79$ ) after applying a cross correlation method on the datasets used [17]. In a different study on maize, [18] MODIS NDVI was used in Zimbabwe to realize strong relationships with the national maize production estimates after the data was adjusted to match onset of the rainy season. The strength of correlations in these two studies is attributed to normalization of the time lag in the climate and NDVI data through the methods used. It is inferred that normalization of satellite data through an appropriate method improves the strength of correlations and is appropriate in future studies. It is also important to note that a combination of satellite and climatic datasets such as those used in these studies utilizes newer methods in forecasting sugarcane productivity [17]. A similar study in Louisiana used thermal variables (Growing Degree Days accumulated from planting to sensing) to adjust in-field NDVI measurements, and to develop a sugarcane yield forecasting method [4]. They obtained a positive exponential correlation, with  $R^2$  improving from  $0.20$ , when using unadjusted NDVI, to  $R^2 = 0.46$ , when using adjusted NDVI. These authors argued that a weak correlation from application of the model was attributed to the spatial variability of sugarcane fields due to different crop ages and diverse environmental conditions in different locations.

In the agricultural landscape of Kenya, sugarcane crop exhibits extreme age differences alongside diversified subsistence cropping in different environmental conditions and is thus highly heterogeneous [19]. MODIS 250 m data has been used successfully to determine temporal dynamics of crops at local scales due to its good geometric and radiometric properties that make the data interoperable with other GIS datasets [20]. However, at MODIS 250 m resolution and in a small agriculture region such as in Kenya, the measured radiation is a mixture of different crops and natural vegetation [19]. It is therefore important to apply a method that will normalize data by removing time lag since this will decrease the effect of mixed crop-natural vegetation pixels in the satellite data used for yield forecasting. The effect of mixed pixels while developing a maize yield model using the land cover weighted NDVI rather than the traditional NDVI reduced the unknown variance by 26% [21]. It was argued that yield estimation using NDVI may vary during respective months of the crop growth because NDVI is reduced at the end of the rainy season, emphasising the need for careful consideration on time integration [11].

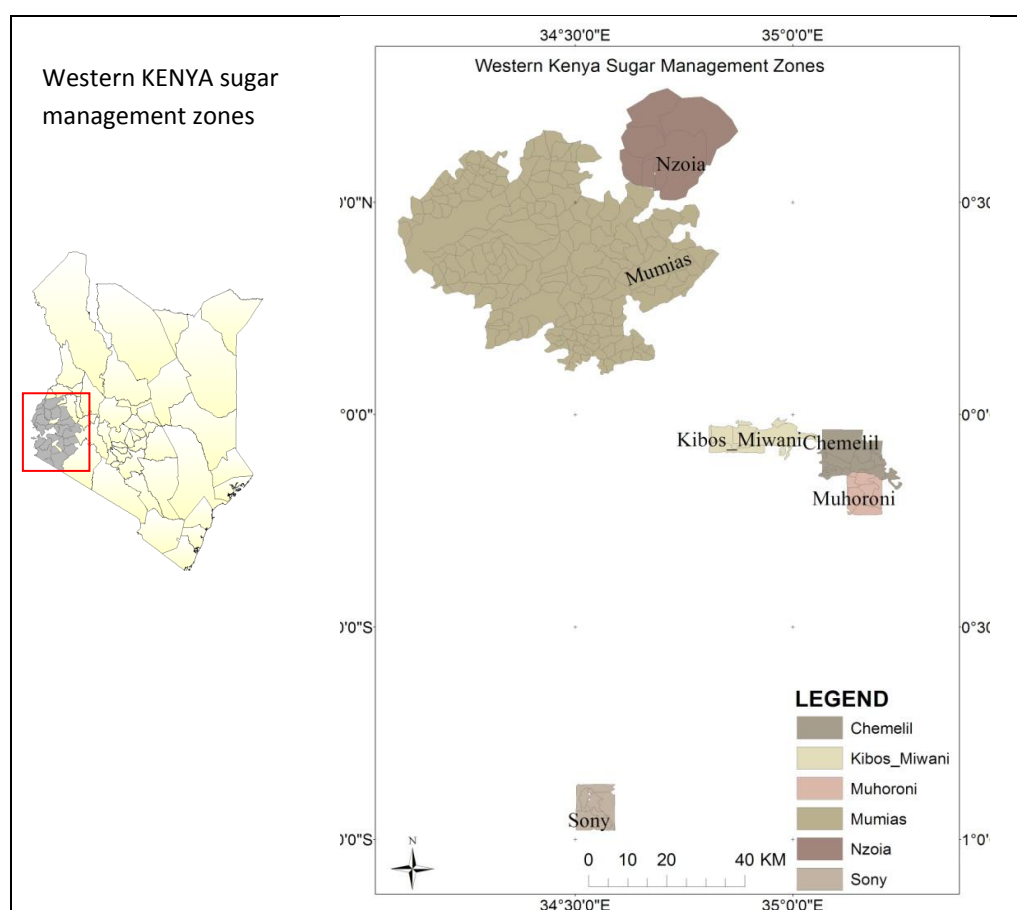
The objective of this study was to test how time integrated Normalized Difference Vegetation Index data from MODIS 250 m imagery can be used for annual sugarcane yield assessment at the sugarcane mill management scales (zones) in Western Kenya. This objective is challenging, since sugarcane in this region is grown in fragmented fields scattered in highly variable environments with various land uses and land covers, soil types, and altitudes. For crop yield forecasting, the ideal approach would be to use crop-specific masks. However, with medium/coarse resolution (about 5–100 ha per pixel) imagery, identifying mono-cropped pixels is not always feasible. This is particularly true in low-producing regions and in regions with sparse crop distribution [14], such as Kenya. Therefore the method proposed here is based on the concept that all vegetation in a region integrates the season's cumulative growing conditions [22]. We first analysed the spatio-temporal variability of the yield-NDVI relationship, using the data set acquired in Western Kenya on six sugarcane zones covering nine years (2002–2010). We used linear models to test the effect of the time integration period of NDVI in relation to the annual yield estimation, and tested the effect of annual rainfall on sugarcane yield. We hypothesize that zones' yield is influenced by cropping practices and environmental conditions at the zonal scale.

## 2. Data and Methods

### 2.1. Study Area

The study area (Figure 1) is located within the western part of Kenya, comprising six sugar management zones that include: (i) Chemelil, Kibos and Muhoroni within the sub humid agro-ecological zone; and (ii) Mumias, Nzoia and Sony within the humid agro ecological zone of Kenya. These zones are located between longitudes 34.18 °E, and 35.87 °E, and latitudes 1.25 °N and 1.50 °S, covering an area of 120,000 ha [23]. Mumias is the highest producer of sugar placed at 39% in 2011 [23]. The landscape of this area is characterized by a mosaic of hills and valleys, with altitudes ranging from 1,000 m (Kibos) to 1,600 m (Mumias and Nzoia), and 1,800 m (Chemelil), and slope rising between 8%, in the plains of Kibos zone, and 38%, in the hills of Chemelil zone.

**Figure 1.** Western Kenya sugar management zones (Source: Sugar mills).



The topography influences the agro-ecological zones receiving an average of 1,400 mm and 1,800 mm of rainfall in the sub humid and humid zones, respectively [24]. Rainfall in this area is bimodal [25] with a long rain season between March and July, with planting in March for food crops and April for sugarcane; and a short rain season in September to December with planting in September for all crops [26]. This variation in rainfall distribution influences an intensified cropping system with crop diversification and rotation of food crops and sugarcane age. Soils of the study area are dominantly black cotton cambisols in the low lands and sandy loamy acrisols in the highlands [27].

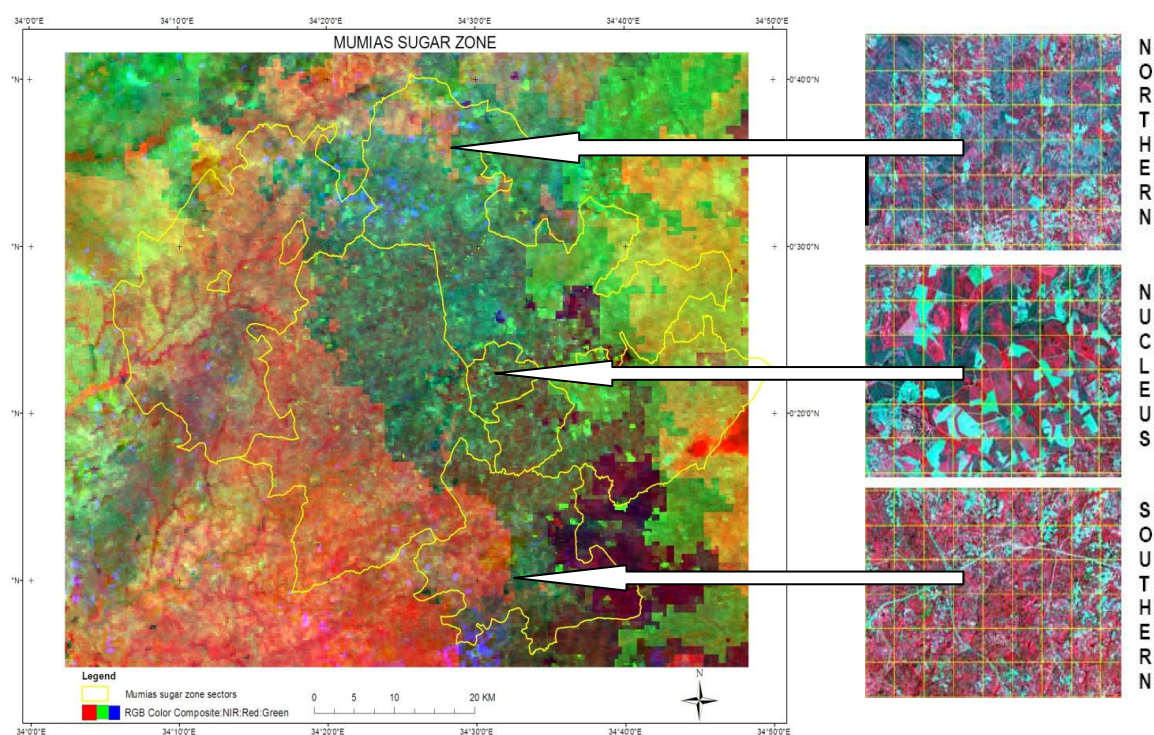
The location of the study area in different agro-climatic zones, diversified topography, soils types and cropping practices provides an ideal scenario to explore the relationship between sugarcane productivity, environmental variables, and management practices in Western Kenya.

## 2.2. Data

### 2.2.1. Satellite Data and Pre-Processing

A complete 11-year time series (2002–2012) of the Surface Reflectance 8-Day L3 Global 250m product (MOD09Q1) was downloaded through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC) [28]. MOD09Q1 product provides bands 1 (red reflectance; 620–670 nm) and 2 (near infrared reflectance; 841–876 nm) at 250-m resolution. Each MOD09Q1 pixel contains the ‘best possible observation’ during an 8-day period as selected on the basis of high observation coverage, low view angle, the absence of clouds or cloud shadow, and aerosol loading. The accuracy of the version-5 MODIS/Terra Surface Reflectance products has been assessed over a widely distributed set of locations and time periods via several ground-truth and validation efforts, and thus ready for use in scientific publications. The red (R) and (NIR) reflectance data were used to compute the NDVI [29] for all the 460 images.

**Figure 2.** (a) MODIS 250 m color composition of Mumias zone (sectors within the zone are delineated by a yellow line), and (b) subsets of a December 2011 SPOT 2.5 m image on three sectors; the overlaying yellow grids correspond to the 250 m spatial resolution of MODIS pixels.



@ LP DAAC online data pool

(a)

@ CNES 2011, Distribution Spot Image

(b)

In addition to the MODIS time series, a multispectral (Green, Red, and Near Infrared) 2.5 m SPOT image was acquired over Mumias in December 2011. This data was used to appraise land cover and use in different sectors of Mumias sugar zone in a 250 m grid (Figure 2), showing the large heterogeneity of the landscape at MODIS scale, and the impossibility to use a sugarcane crop mask on a satellite image at MODIS scale in the area.

### 2.2.2. Agronomic and Climatic Data

The agronomic (yield and cropped area) and climatic data were obtained from the respective sugar mills. At the zonal scale, yearly cropped area (ha), estimated yield ( $\text{tC}\cdot\text{ha}^{-1}$ ), and monthly rainfall data were obtained for the period 2002 to 2010. We also obtained yield data for the year 2012 which was used for quantitative validation of the model. Crop area data are estimated by physical measurement of area that has been harvested or during land preparation. On the other hand, yearly yield is obtained using the Visual Physical Assessment method (as presented in the Introduction section).

Rainfall data were recorded using 113 rain gauges distributed unequally among all the sugar zones. The rainfall data was cross tabulated to compute the annual mean for each zone for comparison with the annual yield.

## 2.3. Data Analysis

### 2.3.1. Time-Integration of NDVI Values

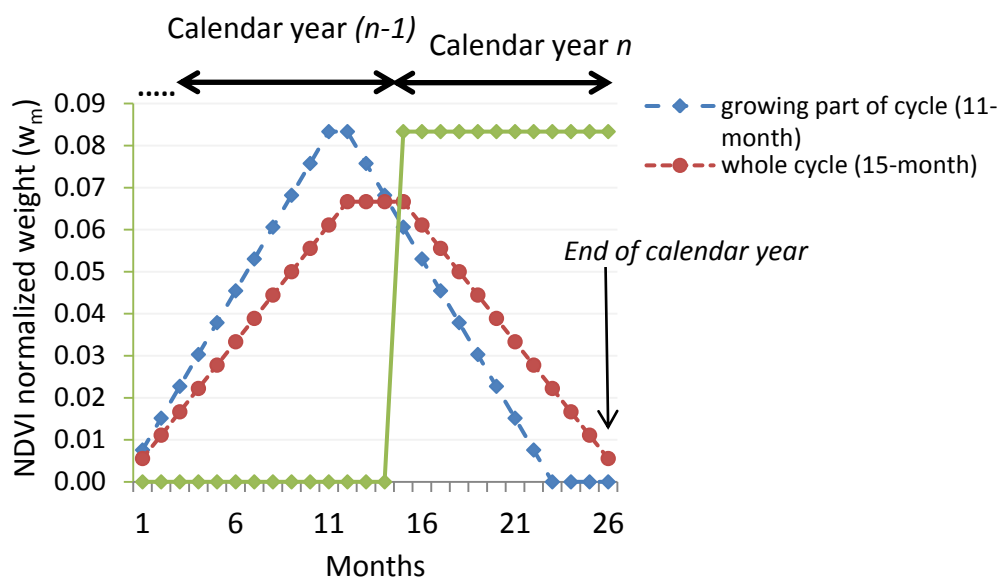
A thematic layer of the limit of the sugarcane growing mill zones was used to extract 8-day NDVI values for each zone. These NDVI values were then spatially aggregated to allow comparison with the mean annual yield, at the same scale. Generally, time integration of NDVI is done throughout the calendar year [2,11,18]. At the field scale, [10,20] is considered a seasonal integration approach which utilized either the sowing or the harvesting date, while at the regional scale, [4] used growing degree days to compute in season NDVI for estimating yield and obtained good results. At regional scale in Portugal, [30] correlated yield of the current year with a 10-day NDVI data to develop a yield estimation model which explained 77%–88% of wine yield. At state scale in Brazil, [31] used thermal time other than the calendar year and estimated sugarcane yield with a RMSE of  $1.5 \text{ t}\cdot\text{ha}^{-1}$  (around 2% of accuracy); however, they used a crop mask and selected sugarcane pixel purity above 95% for the establishment of the regressions.

In this study therefore, we tested a new way of time integration in order to account for the local sugarcane cropping practices at zonal scale. In effect, since the yield is estimated on a calendar year base (harvest lasts from January to December), a ratoon crop growing from November 2009 to its harvest in January 2011—at the age of 15 months—accounts for the 2011 annual yield data. Therefore, this complicates the yield prediction scenario where, in this case, the 2011 annual yield includes the yield of a crop that was almost nonexistent on the 2011 satellite time series (except on the January image). It is argued that predicting yield in such small rain fed sugarcane fields is complicated since NDVI from all land uses declines at the end of the rainfall period [11] and requires a keen consideration of the integration period. In a similar case, a weighted land cover NDVI was used to account for the influence of other land uses on maize yield [21]. We therefore applied a weighting



matrix over a period of time corresponding to the growing calendar, and not to the calendar year in order to take into account the active vegetative stages of the crop and minimize any shift in NDVI during sugarcane development [22]. To do this we chose two different periods of integration, (1) an 11-month period which corresponds to the approximate length of the growing cycle before maturation, and (2) a 15-month period which corresponds to the approximate length of the whole growing cycle. For both configurations, we calculated a weight for each month corresponding to the probability of a sugarcane field to be harvested during the calendar year of yield estimations, and thus to be accounted for in the annual yield (Figure 3).

**Figure 3.** Three sets of weights used to calculate time integration of monthly NDVI values for annual yield estimation (year  $n$ ). The green line (between months 14 to 26) corresponds to weights generally used to calculate the annual NDVI (the calendar year corresponding to the yield measurement). The blue and red lines correspond to weights that take into account the sugarcane cropping calendar (15 months for the whole cycle, and 11 months for the growing period) in the NDVI time integration.



Annual NDVI (NDVI) and weighted NDVI (wNDVI\_15 and wNDVI\_11) for each year was calculated according to Equation (1), with  $i$  equals to 15 and 11, respectively. The value 15 corresponds to the length of the usual cropping cycle of the sugarcane (in months), while the value 11 corresponds to the length of the vegetative part (in months) which is mainly related to cane yield [10].

$$wNDVI\_i = \sum_{m=1}^{m=i} NDVI_m w_m \quad (1)$$

where,  $NDVI_m$  is the value of the NDVI for month  $m$ ,  $w_m$  is a coefficient equal to the NDVI normalized weight (Figure 3), and  $i$  is the length of the time integration (in months). The sum of the  $w_m$  coefficients is equal to 1.



### 2.3.2. Spatio-Temporal Analysis

The relationship between NDVI, wNDVI and the annual estimated yield was studied with a linear regression [20,32] and exponential regressions [4] established through time and space using a one-tailed probability test. We then assessed the role of the environmental variables in the relation between yield and NDVI, by correlating the slope of the “yield-NDVI (wNDVI)” relationships with the rainfall, and with the sugarcane fraction in each respective zone.

## 3. Results and Discussion

### 3.1. Yield and Climatic Data Variability

Table 1 demonstrates variable annual rainfall distribution within the six zones ranging between 1,421 mm and 1,869 mm. This rainfall groups the sugar zones into two climatic regions: the sub humid with less than 1,500 mm (Kibos, Chemelil and Muhoroni) and humid with about 1,800 mm or more (Sony, Mumias and Nzoia) agro-ecological zones (AEZ) respectively, both lying within the tropical climate of the country.

**Table 1.** Summary of the agronomic and climate data used in the study: mean and standard deviation (in parenthesis) calculated over the 9-year period (2002–2010).

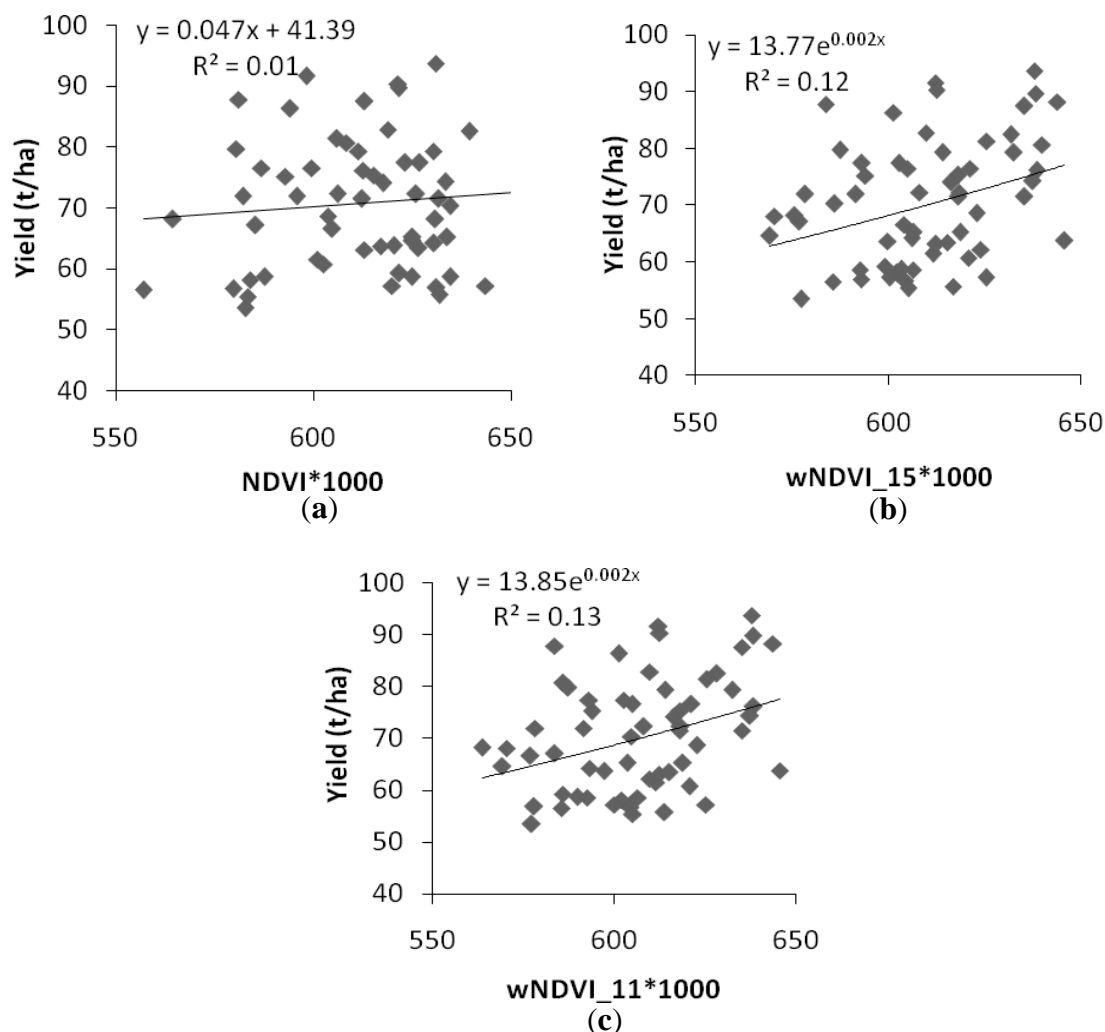
	KIBOS	MUMIAS	CHEMELIL	MUHORONI	SONY	NZOIA
Rainfall (mm yr <sup>-1</sup> )	1,421 (102)	1,835 (186)	1,426 (263)	1,486 (214)	1,869 (221)	1,763 (252)
Yield (t·ha <sup>-1</sup> )	71.1 (9.6)	75.6 (11.1)	62.6 (9.6)	63.9 (7.9)	80.1 (11.3)	75.0 (5.2)
Sugarcane fraction (%)	32.2 (4.5)	48.7 (2.5)	38.8 (6.3)	50.5 (7.3)	33.3 (5.3)	22.2 (2.7)

Sugarcane grown in regions with less than 1,500 mm rainfall is recommended for supplemental irrigation [1]. The reason for higher yield in Kibos (71 t ha<sup>-1</sup>), compared to the government owned Chemelil and Muhoroni sugar mills in the same AEZ whose yield is around 63 t ha<sup>-1</sup> can be explained by better crop husbandry. Globally, yield in the humid AEZ (Mumias, Sony, and Nzoia) is higher (between 75 and 80 t ha<sup>-1</sup>) than in the sub-humid AEZ. The yield in Sony (80 t ha<sup>-1</sup>) is boosted by large scale farmers within the fertile highlands of Transmara.

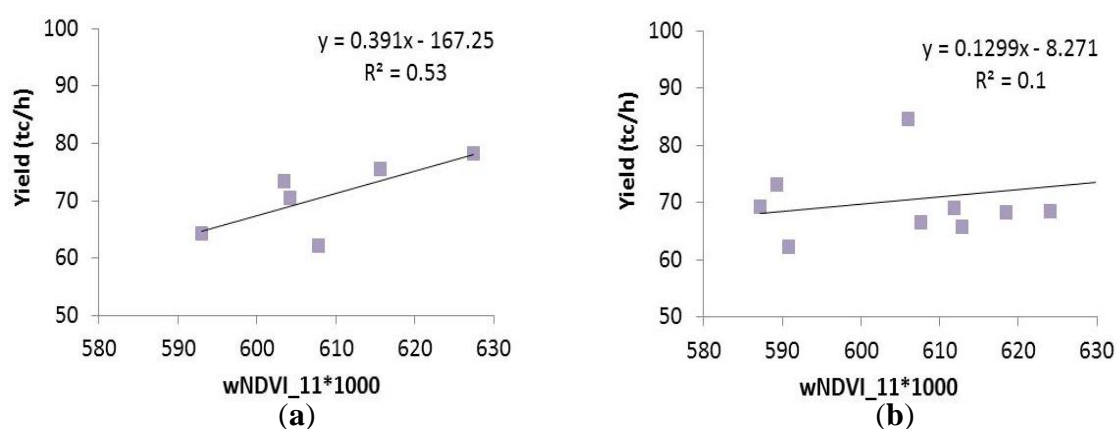
### 3.2. Relationship between Yield and NDVI

When the whole data set (6 zones and 9 years) is used, the analysis shows that the annual NDVI is not strongly related to the sugarcane yield ( $p = 0.1$ ; (Figure 4(a)). This finding is close to those who found no relationship between average NDVI and farmers’ yield [11] and; whose results showed low significance when correlating historical yield and NDVI at annual level ( $P = 0.1$ ) [2]. However, when adjusted NDVI (wNDVI) is used, the relationship is highly significant for wNDVI\_11 ( $P = 0.001$ ) (Figure 4(c)) and significant for wNDVI\_15 ( $P = 0.01$ ) (Figure 4(b)) with the  $R^2$  increasing from 0.01 to 0.12 and 0.13 respectively through both linear and exponential relationships. This result is in agreement with a study demonstrating that yield estimations based on metrics obtained a little after the peak of APAR can be done without seriously compromising performance [31]. However, the strength of these correlations is weak, justifying further analysis by this study on other factors that affect yield.

**Figure 4.** Relationship between (a) yield and annual NDVI, (b) yield and wNDVI\_15, and (c) yield and wNDVI\_11.



**Figure 5.** Variability with wNDVI\_11 averaged (a) at zone level on the 2002–2010 periods, and (b) at annual level on the six zones.



When the whole dataset is aggregated over the whole period (2002–2010), at the zone level (spatial analysis), the correlation between yield and wNDVI is significant (Figure 5(a)) with  $R^2 = 0.53$ ,  $P < 0.001$ ; while when the whole dataset is aggregated over the six zones, at the year level (temporal analysis); there is no significant correlation between yield and wNDVI (Figure 5(b)). The good result

obtained through the spatial analysis is due to different environmental variables exuded through rainfall distribution. The absence of significant results through the temporal analysis could be explained by (1) the difficulty to make coherent yield measurements over a calendar year and wNDVI (considering the length of time sugarcane takes to mature), and (2) the sugarcane cover fraction changes during the 2002–2010 period (see standard deviation values of the fraction of sugarcane cropped area in each zone, Table 1).

### 3.3. Relationship between Yield and Rainfall

In order to better understand the spatial and temporal variability of yield, we studied the relationship between yield and rainfall. When using all the data (6 zones  $\times$  9 years; Figure 6(a)), the relation between annual yield and rainfall was significant, but weak ( $R^2 = 0.08$ ;  $p = 0.03$ ). Such a weak relationship has been attributed to the time lag between yield and rainfall because vegetation takes a considerable period to respond to soil moisture [25]. This effect is amplified in Western Kenya, where the annual yield is dependent on the rainfall of the previous year due to the length of the sugarcane cycle. On removal of the time lag through spatial and temporal averaging over the nine year data (6 zones  $\times$  9 years; Figure 6(b,c)), this study showed a strong relationship as noted by other studies [4,25] with  $R^2 = 0.8$  and  $p < 0.001$  at the spatial level (Figure 6(b)). The relationship between yield and rainfall (Figure 6(b)) is stronger than the relationship between yield and wNDVI (Figure 5(a)) at the zone scale. This is because unlike rainfall which is an environmental variable, wNDVI value integrates not only sugarcane area, but also other types of land covers that are in different proportions according to the zone. The temporal analysis of yield and rainfall shows no correlation between both variables (Figure 6(c)), because (1) rainfall is not the only yield driving factor, and (2) because annual rainfall should be integrated on a longer period and with different weights (as wNDVI) in order to take into account the particular cropping calendar of the sugarcane crop. These results are in agreement with a study that pointed out that rainfall amounts and pattern may not be a reliable predictor of yield [11].

**Figure 6.** Relationship between yield and rainfall using: (a) all the data, (b) the data aggregated at the zone scale (spatial analysis), and (c) the data aggregated at annual scale (temporal analysis).

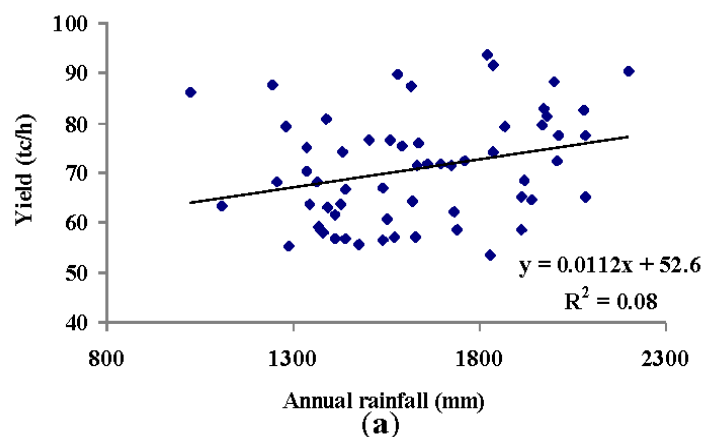
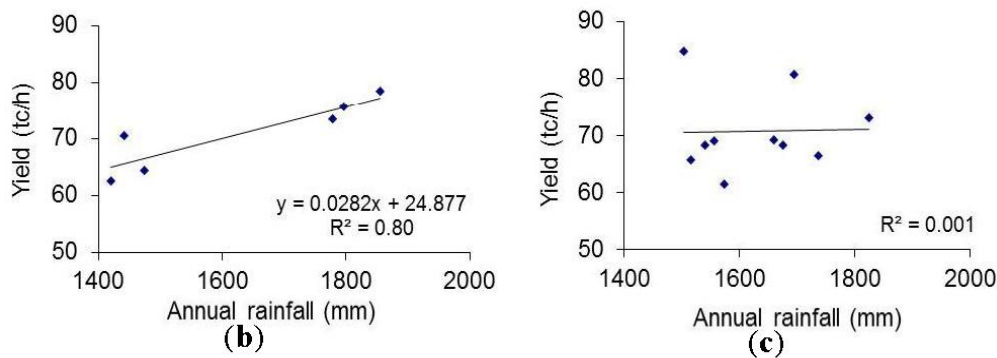


Figure 6. Cont.



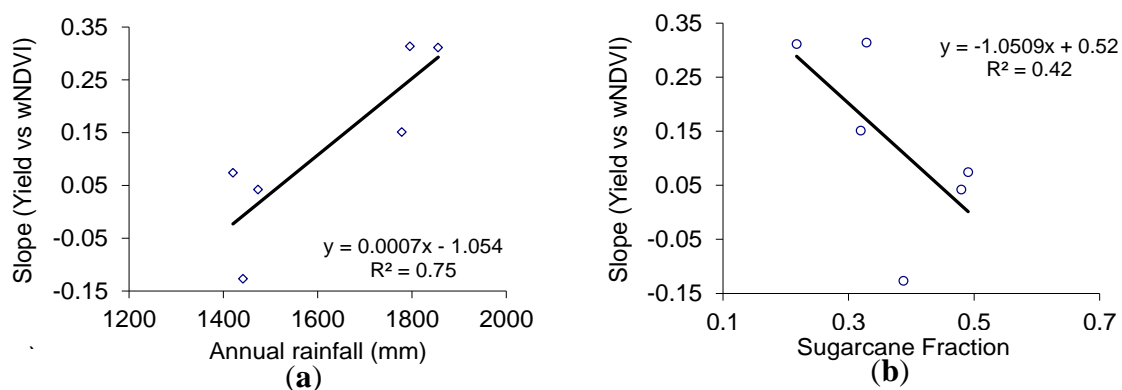
### 3.4. Relationship between Yield-wNDVI Slope and Rainfall

In order to better understand the main driving factors of the yield-wNDVI relationship, we correlated the slope of the relation between yield and wNDVI aggregated at the zone scale with the rainfall (Figure 7(a)), and with the fraction of sugarcane in each zone (Figure 7(b)). Results show a strong correlation with high significance at  $p < 0.001$  in both cases.

The sensitivity of the yield-wNDVI variations to each millimeter rainfall received in each management zone also called the Precipitation Marginal Response, or PMR [33], separates two groups of three zones geographically located in sub humid AEZ from those in the humid AEZ (Figure 7(a)). The ability to separate the two climatic regimes in this study therefore strengthens the ability to use wNDVI in forecasting crop yield. Results of this relationship were highly significant with  $R^2 = 0.75$ ;  $P = 0.001$ . The positive slope of this relationship (Figure 7(a)) indicates that the sensitivity of the yield to rainfall is higher than the sensitivity of the wNDVI to rainfall.

The negative slope resulting from the relationship between yield-wNDVI slope and sugarcane fraction (Figure 7(b)) indicates that wNDVI is not only affected by the amount of rainfall received in the zone, but is also influenced from other surrounding vegetation cover [2] considering that sugarcane has larger biomass than the surrounding environment.

**Figure 7.** Relationship between the “yield-wNDVI” slope and (a) rainfall, and (b) sugarcane fraction, aggregated at the zone scale.



### 3.5. A Quantitative Evaluation of the Model

WNDVI\_11 data for the year 2011 and 2012 was used to estimate the 2012 sugarcane yield (Table 2) using the model established at the zone scale (Figure 5(a)), in order to utilize data that is independent from the one used in development of this model.

**Table 2.** Model validation using 2012 yield.

Zone	wNDVI_11	Model Yield (t ha <sup>-1</sup> )	Measured Yield (t·ha <sup>-1</sup> )	Squared Error (t ha <sup>-1</sup> )
Mumias	566.5	54.2	48	38.44
Nzoia	602.8	68.4	64.7	13.69
Chemelil	586.9	62.2	59	10.24
Muhoroni	604.4	69.1	63.6	30.25
Kibos	596.1	65.8	62.7	9.61
Sony	610.5	71.5	69	6.25
RMSE				4.25

We obtained a Root Mean Squared Error (RMSE) of 4.25 t ha<sup>-1</sup> when all the zones are considered. The worst yield estimation was realized in Mumias zone (+6.2 t ha<sup>-1</sup>), where the land holdings are particularly small (up to 0.1 ha), and where the landscape is very heterogeneous (Figure 2). This result is similar to the low accuracy obtained for fields smaller than the pixel size and high accuracies for large fields [15]. When excluding Mumias zone, the RMSE decreases to 3.41 t ha<sup>-1</sup>, which is in agreement in both cases, with the user specification of RMSE 5 t ha<sup>-1</sup>.

## 4. General Discussion and Conclusions

This research has investigated the influence of cropping practices and environmental conditions on yield at zone scale through two approaches. Firstly, historical yield was related to annual NDVI with the assumption that yearly sugarcane yield is significantly correlated to annual NDVI. This hypothesis was rejected since the significance of this correlation was only achieved after adjusting the NDVI through time integration of the sugarcane growing period to remove the time lag in crop growth. The strength of this relationship was then enhanced when the data were aggregated over the whole period (2002–2010) at the zone level. Secondly, historical yield was related to rainfall and the strength of this relationship was low, although the correlation was of high significance. The relationship was equally strengthened through spatial aggregation and through rain use efficiency. The relation between yield and rainfall exists owing to the fact that sugarcane yield is significantly related to rainfall on removal of time lag at zone scale since crops take a considerable period to respond to rainfall.

This study has shown that remote sensing technology together with environmental information has potential to be used to estimate crop yield and evaluate the impact of environmental conditions to crop production as opposed to physical methods. In effect, it has been reported that accuracy of physical methods such as visual physical approach (VPA) on yield estimation is minimized due to gross errors associated with fatigue, variability in assessment of natural phenomena using the naked human eye, and lack of consideration of diverse environmental variables (such as rainfall) during the growth period of the cane crop [6]. The use of remote sensing data can highlight variations in environmental variables within respective zones, and this is uniquely evidenced by the separation of the two

agro-ecological zones through spatial aggregation. Additionally, variations within and between the zones are influenced by environmental variables such as soil characteristics and rainfall distribution over different years. Our findings are in agreement with a study noting that rainfall was not the single determinant of crop yield in different environments, but rather, other factors such as soil characteristics, and other agricultural land use need to be included [8].

In summary, our results are in agreement with most of the previous studies on this subject. Through this study, we have contributed knowledge to remote sensing fraternity (1) by developing an original method for NDVI time integration that takes into account the local cropping practices (length of the growing season), and (2) by analyzing the spatial and temporal dimensions of the yield-NDVI relationship and response of its slope to rainfall. Sugarcane yield forecasting has been exemplified through spatial aggregation of weighted NDVI. The information presented in this study is useful for proper, foresighted and informed planning in the Kenya's Sugar Industry at the zone management scale. This is because the information explains the influence of environmental conditions on sugarcane production, thus providing knowledge for monitoring sugarcane productivity at the zone scale.

Further research is recommended by this study, to refine the zone scale to farm level. The issue of scale is suggested to minimize the influence of other land cover on NDVI extracted from sugarcane fields. Future Earth Observing satellite systems, such as Sentinel-2 (ESA), with decametric spatial resolution, and a high visiting frequency, will give access to farm level information.

## Acknowledgements

The authors particularly appreciate CIRAD and KESREF for sponsoring this research. Much thanks also to the USGS based GLOVIS data repository for providing free access to MODIS time series data. Special thanks to the sugar mills: Mumias, Kibos-Miwani, Chemelil, Nzoia and Sony, for providing the yield and rainfall data, and availing field assistants during field work.

## Conflict of Interest

The authors declare no conflict of interest.

## References

1. Kenya Sugar Research Foundation (KESREF). *Sugarcane Growers' Guide*; Kenya Sugar Research Foundation: Kisumu, Kenya, 2010.
2. Bastidas-Obando, E.; Carbonell-Gonzalez, J. Evaluating the Applicability of MODIS Data for Forecasting Sugarcane Yields in Colombia. In Proceedings of International Society of Sugar Cane Technologists (ISSCT), Durban, South Africa, 29 July–2 August 2007.
3. Kenya Sugar Board. Cane Census 2009/2010–2010/2011. In *Agriculture*; Kenya Sugar Board: Nairobi, Kenya, 2009.
4. Lofton, J.; Tubana, B.S.; Kanke, Y.; Teboh, J.; Viator, H.; Dalen, M. Estimating sugarcane yield potential using an in-season determination of normalized difference vegetative index. *Sensors* **2012**, *12*, 7529–7547.

5. Lamb, J.A.; Dowdy, R.H.; Anderson, J.L.; Rehm, G.W. Spatial and temporal stability of corn grain yields. *J. Prod. Agr.* **1997**, *10*, 410–414.
6. Longley, P.A.; Goodchild, M.F.; Maguire, D.J.; Rhind, D.W. *Geographical Information Systems (Abridged): Principles, Techniques, Management, and Applications*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005; p. 358.
7. Pettorelli, N.; Vik, J.O.; Mysterud, A.; Gaillard, J.M.; Tucker, C.J.; Stenseth, N.C. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol.* **2005**, *20*, 503–510.
8. Zarco-Tejada, P.J.; Ustin, S.L.; Whiting, M.L. Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. *Agron. J.* **2005**, *97*, 641–653.
9. El Hajj, M.; B égu é A.; Guillaume, S.; Martin é J.-F. Integrating SPOT-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices—The case of sugarcane harvest on Reunion Island. *Remote Sens. Environ.* **2009**, *113*, 2052–2061.
10. B égu é A.; Lebourgeois, V.; Bappel, E.; Todoroff, P.; Pellegrino, A.; Baillarin, F.; Siegmund, B. Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. *Int. J. Remote Sens.* **2010**, *31*, 5391–5407.
11. Gunnula, W.; Kositrakun, M.; Righetti, T.; Weerathaworn, P.; Prabpan, M. Normalized difference vegetation index relationships with rainfall patterns and yield in small plantings of rain-fed sugarcane. *Aust. J. Crop Sci.* **2011**, *5*, 1845–1851.
12. Goward, S.N.; Dye, D.G. Evaluation North American net primary productivity with satellite observations. *Adv. Space Res.* **1987**, *7*, 165–174.
13. Prince, S.D. Satellite remote sensing of primary production: Comparison of results for Sahellian grassland 1981–1988. *Int. J. Remote Sens.* **1991**, *12*, 1301–1312.
14. Atzberger, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens.* **2013**, *5*, 949–981.
15. Fernandes, J.L.; Rocha, J.V.; Lamparelli, R.A.C. Sugarcane yield estimates using time series analysis of SPOT Vegetation images. *Sci. Agr.* **2011**, *6*, 139–146.
16. Bastiaanssen, W.G.M.; Ali, S. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. *Agr. Ecosyst. Environ.* **2003**, *94*, 321–340.
17. Gonçalves, R.R.V.; Zullo, J., Jr; Romani, L.A.S.; Nascimento, C.R.; Traina, A.J.M. Analysis of NDVI time series using cross-correlation and forecasting methods for monitoring sugarcane fields in Brazil. *Int. J. Remote Sens.* **2012**, *33*, 4653–4672.
18. Funk, C.; Budde, E.M. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sens. Environ.* **2009**, *113*, 115–125.
19. Mulianga, B.; B égu é A.; Simoes, M.; Todoroff P.; Clouvel, P. MODIS Data for Forecasting Sugarcane Yield in Kenya through a Zonal Approach, In Proceedings of the Sentinel-2 Preparatory Symposium, Frascati (IT), Frascati, Italy, 23–27 April 2012.
20. Nguyen, T.T.H. Spatial and temporal unmixing of coarse resolution polar orbiter observations for sugarcane yield assessments—A seasonal integration method based on MODIS and MISR FPAR estimates in Xinavane, Mozambique. M.Sc. Thesis, International Institute for Geo-Information Science and Earth Observation: Enschede, NL, USA, 2005; p. 50.



21. Rojas, O. Operational maize yield model development and validation based on remote sensing and agrometeorological data in Kenya. *Int. J. Remote Sens.* **2007**, *28*, 3775–3793.
22. Kastens, J.H.; Kastens, T.L.; Kastens, D.L.A.; Price, K.P.; Martinko, E.A.; Lee, R.-Y. Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sens. Environ.* **2005**, *99*, 341–356.
23. Kenya Sugar Board. Year Book of Statistics. In *Sugar Board Papers*; Kenya Sugar Board: Nairobi, Kenya, 2010.
24. Ribot, J.C.; Magalhães, A.R.; Panagides, S.S. *Climate Variability, Climate Change and Social Vulnerability in the Semi-Arid Tropics*; Cambridge University Press: Cambridge, UK, 2005.
25. Shisanya, C.; Recha, C.; Anyamba, A. Rainfall variability and its impact on normalized difference vegetation index in arid and semi arid lands of Kenya. *Int. J. Geosci.* **2011**, *2*, 36–47.
26. Amolo, R.; Abayo, G.; Muturi, S.; Rono, J. *The Impact of Planting and Harvesting Time on Sugarcane Productivity in Kenyan Sugar Industry*; KESREF: Kisumu, Kenya, 2009; Volume 3, pp. 18–27.
27. Jaetzold, R.; Schmidt, H.; Hornetz, B.; Shisanya, C. *Farm Management Handbook of Kenya—Natural Conditions and Farm Management Information*; Ministry of Agriculture, Farm Management Section: Nairobi, Kenya, 1985; Volume 2, p. 317.
28. United States Geological Survey (USGS)-Global Visualization Viewer (GloVis). NASA Land Processes Distributed Active Archive Center (LP DAAC). Online Data Pool. Available online: [https://lpdaac.usgs.gov/get\\_data](https://lpdaac.usgs.gov/get_data) (accessed on 7 December 2012).
29. Rouse, J.W.; Hass, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, Washington, DC, USA, 10–14 December 1973; pp. 309–317.
30. Cunha, M.; Andre, R.S.M.; Lisa, S. Very early prediction of wine yield based on satellite data from vegetation. *Int. J. Remote Sens.* **2010**, *31*, 3125–3142.
31. Duveiller, G.; López, L.R.; Baruth, B. Enhanced processing of 1-km spatial resolution fAPAR time series for sugarcane yield forecasting and monitoring. *Remote Sens.* **2013**, *5*, 1091–1116.
32. Rasmussen, M.S. Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR. *Int. J. Remote Sens.* **1992**, *13*, 3431–3442.
33. Veron, S.R.; Oosterheld, M.; Paruelo, J.M. Production as a function of resource availability: Slopes and efficiencies are different. *J. Veg. Sci.* **2005**, *16*, 351–354.